# Exploiting LLMs for metadata-based video quality prediction

Steve Göring, Rakesh Rao Ramachandra Rao, Alexander Raake[+]

Audiovisual Technology Group,
Technische Universität Ilmenau, Germany;
Email: [steve.goering, rakesh-rao.ramachandra-rao]@tu-ilmenau.de
[+] Chair of Signal Processing in Communication Engineering, IENT; RWTH Aachen, Germany; raake@ient.rwth-aachen.de

Code&Data: `https://bit.ly/3KKYNNr`

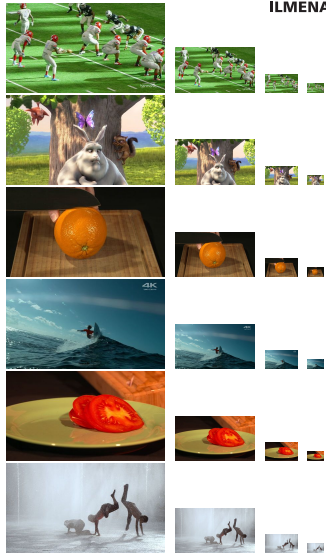
December 8, 2025

TECHNISCHE UNIVERSITÄT
ILMENAU

# Motivation

TECHNISCHE UNIVERSITÄT
ILMENAU

- most video streaming DASH/HAS
  - several representations
    - framerate, codec, resolution, bitrate
  - quality prediction required (e.g. playout adaption)
  - e.g., AVQBits|M0 (metadata) [2], P.1203.3 [3, 6],
  - ITU-T P.1204.3 (bitstream) [5]
- LLMs gaining polularity in various fields
- **LLMs for meta-data based quality prediction**



[4]

# Approach



- prompt template: """let's assume you watch on a 4K screen a video encoded with {codec} using a {bitrate} bitrate, {fps} fps, and a {resolution} resolution, what would be the perceived quality using a 1=bad to 5=best continuous scale, answer only the score in a json format with key score"""
- used LLMs: 17 models; Ollama or APIs
  - *llama3.3*, *llava*, *mistral*, *phi4*, *qwen2.5*,
  - *deepseek-r1-:{1.5b, 7b, 8b, 14b, 32b, 70b, web}*, *deepseek-v3-:web*,
  - *gemini-2.0-{flash, flash-light}*, *chatgpt-o4-mini*, and *chatgpt-o1-preview*.
- dataset: test #1 AVT-VQDB-UHD-1 dataset [4]
  - 180 different encoded videos (6 sources, 3 codecs, 4 resolutions, various bitrates)

# Evaluation

TECHNISCHE UNIVERSITÄT
ILMENAU

| Model | Type | Pearson | Kendall | Spearman | *rmse* |
|---|---|---|---|---|---|
| *bestsubset* | *LLM* | 0.902 | 0.720 | 0.897 | 0.523 |
| *meanmodel-no-out* | *LLM* | 0.895 | 0.706 | 0.887 | 0.534 |
| *meanmodel* | *LLM* | 0.878 | 0.689 | 0.877 | 0.612 |
| *deepseek-v3:web* | *LLM* | 0.870 | 0.734 | 0.877 | 0.569 |
| *gemini-2.0-flash-light* | *LLM* | 0.857 | 0.710 | 0.860 | 0.586 |
| *gemini-2.0-flash* | *LLM* | 0.810 | 0.673 | 0.826 | 0.681 |
| *deepseek-r1:web* | *LLM* | 0.808 | 0.654 | 0.833 | 0.900 |
| *deepseek-r1:32b* | *LLM* | 0.786 | 0.635 | 0.789 | 0.823 |
| *P.1204.3* [2] | *BS* | 0.968 | 0.822 | 0.953 | 0.280 |
| *AVQBits\|M0* [2] | *M* | 0.891 | 0.703 | 0.888 | 0.507 |
| *AVQBits\|H0\|s* [2] | *M+NR* | 0.964 | 0.817 | 0.954 | 0.298 |
| *AVQBits\|H0\|f* [2] | *M+NR* | 0.957 | 0.805 | 0.946 | 0.324 |

- considering LLMs as individual raters: LLMs $\approx$ *human* (subject bias, inconsistency based on ITU-T P.910 Annex E; SOS-Analysis [1])

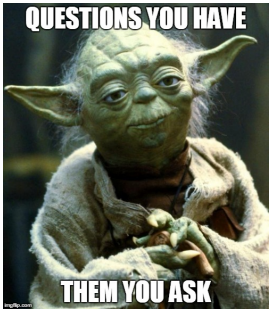
# Conclusion, Summary and Future Work



- core idea and results
  - using LLMs for meta-data based video quality prediction
  - larger models perform better, e.g. *deepseek-v3:web*, or combiantions *bestsubset*
  - rather a proof of concept, becasue LLMs are slow
- open source dataset, evaluation & comparison
  - evaluation of 17 models
  - as ensemble (using mean, and other aggregations)
  - as individual raters
- future work: LLMs for:
  - additional validation scores
  - the design of quality assessment tests

# Thank you for your attention






. . . . . are there any questions?

We want to thank the "**AG Wissenschaftliches Rechnen**" of the TU Ilmenau for providing computing resources.

# References I



[1] Tobias Hoßfeld, Raimund Schatz, and Sebastian Egger. "SOS: The MOS is not enough!" In: *2011 third international workshop on quality of multimedia experience*. IEEE. 2011, pp. 131–136.

[2] Rakesh Rao Ramachandra Rao, Steve Göring, and Alexander Raake. "AVQBits-Adaptive Video Quality Model Based on Bitstream Information for Various Video Applications". In: *IEEE Access* 10 (2022).

[3] Rakesh Rao Ramachandra Rao et al. "Adaptive video streaming with current codecs and formats: Extensions to parametric video quality model ITU-T P. 1203". In: *EI* (2019).

[4] Rakesh Rao Ramachandra Rao et al. "AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1". In: *IEEE ISM*. Dec. 2019, pp. 1–8.

# References II



[5] Rakesh Rao Ramachandra Rao et al. “Bitstream-based Model Standard for 4K/UHD: ITU-T P.1204.3 – Model Details, Evaluation, Analysis and Open Source Implementation”. In: *QoMEX*. 2020.

[6] Werner Robitza et al. “HTTP Adaptive Streaming QoE Estimation with ITU-T Rec. P.1203 – Open Databases and Software”. In: *ACM MMsys*. 2018.